(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
6 September 2013 (06.09.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/130263 A1

(51) International Patent Classification:
H04N 7/24 (2011.01)

(21) International Application Number:
PCT/US2013/025800

(22) International Filing Date:
13 February 2013 (13.02.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
13/407,759 29 February 2012 (29.02.2012) US

(71) Applicant *(for all designated States except US)*: **MICROSOFT CORPORATION** [US/US]; One Microsoft Way, Redmond, WA 98052-6399 (US).

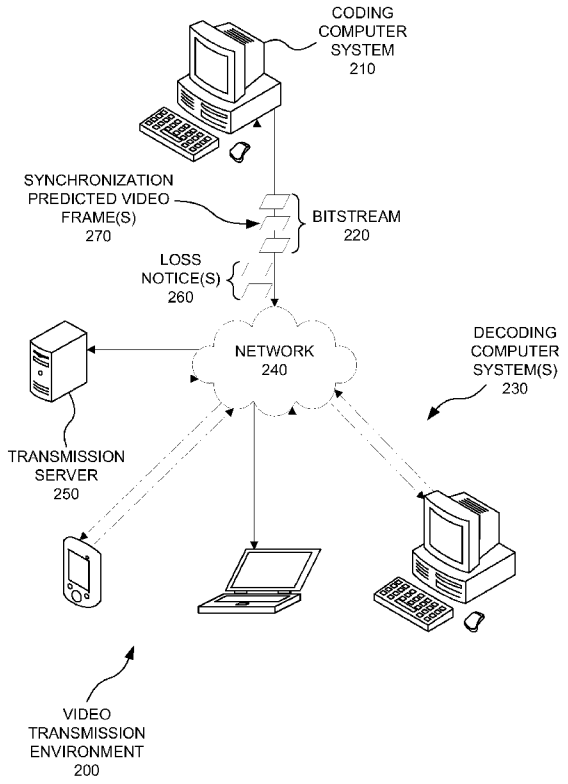
(72) Inventors: **LU, Mei-Hsuan**; c/o Microsoft Corporation, LCA - International Patents, One Microsoft Way, Redmond, WA 98052-6399 (US). **LEE, Ming-Chieh**; c/o Microsoft Corporation, LCA - International Patents, One Microsoft Way, Redmond, WA 98052-6399 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: DYNAMIC INSERTION OF SYNCHRONIZATION PREDICTED VIDEO FRAMES



(57) Abstract: A video bitstream can be encoded and sent over a computer network to a decoding computer system. The bitstream can follow a regular prediction structure when an encoding computer system is not notified of lost data from the bitstream. A notification of lost data in the bitstream can be received. The lost data can include at least a portion of a reference frame of the bitstream. In response, a synchronization predicted frame can be dynamically encoded with a prediction that references one or more other previously-sent frames in the bitstream and that does not reference the lost data. The synchronization predicted frame can be inserted in the bitstream in a position where the regular prediction structure would have dictated inserting a different predicted frame with a prediction that would have referenced the lost data according to the regular prediction structure.

WO 2013/130263 A1

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

— *with international search report (Art. 21(3))*

DYNAMIC INSERTION OF SYNCHRONIZATION PREDICTED VIDEO FRAMES

BACKGROUND

[0001] Predictive coding of video data can improve coding efficiency. However, predictive coding can cause "drift" when some video data is lost in transmission (such as by not arriving at all or arriving too late). "Drift" refers to the propagation of errors from missing data in subsequent frames. For example, when a first video frame (sometimes referred to as a picture) is lost, a second frame that follows the first frame may be coded using prediction that references that first frame. Accordingly, the decoding computer system may be unable to correctly decode that second frame. A third frame may be coded using prediction that references that second frame, and so forth. Indeed, the error from the lost frame (i.e., a frame where at least a portion of the data for the frame was lost) may get worse as subsequent frames are decoded, due to the reliance of the predictive coding on the lost frame. In a conferencing system, intra-coded frames may be inserted in the bitstream to combat this drift problem. For example, intra-coded frames may be periodically inserted in the bitstream. As another example, a coding computer system may dynamically insert an intra-coded frame when the encoding computer system is informed that data from the bitstream has been lost.

SUMMARY

[0002] The disclosure relates to dynamically inserting synchronization predicted video frames. As used herein, dynamically inserted synchronization video frames are video frames that are inserted dynamically and avoid having predictions that rely on specified data, such as lost data. Because these dynamically inserted frames can be predictively coded with reference to previous frames, the frames may be more efficient than comparable intra-coded frames. However, the synchronization predicted video frames can allow for synchronization to cut off drift by avoiding predictions that reference lost data.

[0003] In one embodiment, the tools and techniques can include an encoding computer system encoding and sending a video bitstream over a computer network to a decoding computer system. The bitstream can follow a regular prediction structure when the encoding computer system is not notified of lost data from the bitstream. The encoding computer system can receive a notification of lost data in the bitstream. The lost data can include at least a portion of a reference frame of the bitstream. Also, the encoding computer system can respond to the notification by dynamically encoding a synchronization predicted frame with a prediction that references one or more other

previously-sent frames in the bitstream and that does not reference the lost data. The encoding computer system can insert the synchronization predicted frame in the bitstream in a position where the regular prediction structure would have dictated inserting a different predicted frame with a prediction that would have referenced the lost data according to the regular prediction structure.

[0004] This Summary is provided to introduce a selection of concepts in a simplified form. The concepts are further described below in the Detailed Description. This Summary is not intended to identify key features or essential features of the claimed subject matter, nor is it intended to be used to limit the scope of the claimed subject matter. Similarly, the invention is not limited to implementations that address the particular techniques, tools, environments, disadvantages, or advantages discussed in the Background, the Detailed Description, or the attached drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0005] Fig. 1 is a block diagram of a suitable computing environment in which one or more of the described embodiments may be implemented.

[0006] Fig. 2 is a schematic diagram of a video transmission environment.

[0007] Fig. 3 is a schematic diagram of an example of a regular prediction structure with periodic key frames (top) and a corresponding prediction structure with dynamic key frames (bottom).

[0008] Fig. 4 is a schematic diagram of an example of a regular prediction structure without periodic key frames (top) and a corresponding prediction structure with dynamic long term predicted key frames (bottom).

[0009] Fig. 5 is a schematic diagram of an example of a regular prediction structure without predicted key frames (top) and a corresponding prediction structure with a dynamic anchor predicted frame (bottom).

[0010] Fig. 6 is a flowchart of a technique for dynamic insertion of synchronization predicted video frames.

[0011] Fig. 7 is a flowchart of another technique for dynamic insertion of synchronization predicted video frames.

[0012] Fig. 8 is a flowchart of yet another technique for dynamic insertion of synchronization predicted video frames.

DETAILED DESCRIPTION

[0013] Embodiments described herein are directed to techniques and tools for improved encoding of video bitstreams when a coding computer system is informed that

data from a bitstream has been lost. Such improvements may result from the use of various techniques and tools separately or in combination.

[0014] Such techniques and tools may include dynamically inserting different types of synchronization predicted video frames for different types of regular prediction structures. For example, in a bitstream with periodic key frames, which have predictions that are limited to referring to other key frames, a key frame can be dynamically inserted when a coding computer system is notified of lost data from the bitstream. As another example, in a bitstream without periodic key frames but that allows long term reference key frames, a long term reference key frame can be dynamically inserted when a coding computer system is notified of lost data from the bitstream. Long term reference key frames are key frames that are kept in an active frame window (the window of frames that are to be kept in a decoder frame buffer) for longer than regular frames. For example, a long term reference key frame may be kept in the active frame window until a coding computer system sends an explicit notification to remove that key frame from the active frame window. As another example, in a bitstream with a base layer and an enhancement layer, the base layer of a frame can be coded using prediction that references a previous frame's base layer but does not reference the previous frame's enhancement layer, and the enhancement layer of a frame can be coded using prediction that references a previous frame's enhancement layer but does not reference the previous frame's base layer. With this regular prediction structure, when a coding computing system is informed that data from a frame's enhancement layer has been lost but data from the frame's base layer has not been lost, the coding computing system can dynamically insert an anchor frame. As used herein, an anchor frame is a frame where the base layer is predictively coded with prediction that references a previous frame's base layer and an enhancement layer is intra-coded so that the enhancement layer only references other layers (e.g., the base layer) within the reference frame.

[0015] Accordingly, one or more benefits may be realized from the tools and techniques described herein. For example, the dynamic insertion of synchronization predicted video frames can allow for synchronization to cut off drift, while preserving some efficiency by using some predictive coding where the prediction references data from previous frames.

[0016] The subject matter defined in the appended claims is not necessarily limited to the benefits described herein. A particular implementation of the invention may provide all, some, or none of the benefits described herein. Although operations for the various

techniques are described herein in a particular, sequential order for the sake of presentation, it should be understood that this manner of description encompasses rearrangements in the order of operations, unless a particular ordering is required. For example, operations described sequentially may in some cases be rearranged or performed concurrently. Moreover, for the sake of simplicity, flowcharts may not show the various ways in which particular techniques can be used in conjunction with other techniques.

[0017] Techniques described herein may be used with one or more of the systems described herein and/or with one or more other systems. For example, the various procedures described herein may be implemented with hardware or software, or a combination of both. For example, dedicated hardware implementations, such as application specific integrated circuits, programmable logic arrays and other hardware devices, can be constructed to implement at least a portion of one or more of the techniques described herein. Applications that may include the apparatus and systems of various embodiments can broadly include a variety of electronic and computer systems. Techniques may be implemented using two or more specific interconnected hardware modules or devices with related control and data signals that can be communicated between and through the modules, or as portions of an application-specific integrated circuit. Additionally, the techniques described herein may be implemented by software programs executable by a computer system. As an example, implementations can include distributed processing, component/object distributed processing, and parallel processing. Moreover, virtual computer system processing can be constructed to implement one or more of the techniques or functionality, as described herein.

I. Exemplary Computing Environment

[0018] Fig. 1 illustrates a generalized example of a suitable computing environment (100) in which one or more of the described embodiments may be implemented. For example, one or more such computing environments can be used as a coding computer environment, a decoding computer environment, and/or a server that facilitates transmission of a video bitstream between the coding computer system and one or more decoding computer systems. Generally, various different general purpose or special purpose computing system configurations can be used. Examples of well-known computing system configurations that may be suitable for use with the tools and techniques described herein include, but are not limited to, server farms and server clusters, personal computers, server computers, hand-held or laptop devices, slate devices, multiprocessor systems, microprocessor-based systems, programmable consumer

electronics, network PCs, minicomputers, mainframe computers, distributed computing environments that include any of the above systems or devices, and the like.

[0019] The computing environment (100) is not intended to suggest any limitation as to scope of use or functionality of the invention, as the present invention may be implemented in diverse general-purpose or special-purpose computing environments.

[0020] With reference to Fig. 1, the computing environment (100) includes at least one processing unit or processor (110) and memory (120). In Fig. 1, this most basic configuration (130) is included within a dashed line. The processing unit (110) executes computer-executable instructions and may be a real or a virtual processor. In a multi-processing system, multiple processing units execute computer-executable instructions to increase processing power. The memory (120) may be volatile memory (*e.g.*, registers, cache, RAM), non-volatile memory (*e.g.*, ROM, EEPROM, flash memory), or some combination of the two. The memory (120) stores software (180) implementing dynamic insertion of synchronization predicted video frames.

[0021] Although the various blocks of Fig. 1 are shown with lines for the sake of clarity, in reality, delineating various components is not so clear and, metaphorically, the lines of Fig. 1 and the other figures discussed below would more accurately be grey and blurred. For example, one may consider a presentation component such as a display device to be an I/O component. Also, processors have memory. The inventors hereof recognize that such is the nature of the art and reiterate that the diagram of Fig. 1 is merely illustrative of an exemplary computing device that can be used in connection with one or more embodiments of the present invention. Distinction is not made between such categories as "workstation," "server," "laptop," "handheld device," etc., as all are contemplated within the scope of Fig. 1 and reference to "computer," "computing environment," or "computing device."

[0022] A computing environment (100) may have additional features. In Fig. 1, the computing environment (100) includes storage (140), one or more input devices (150), one or more output devices (160), and one or more communication connections (170). An interconnection mechanism (not shown) such as a bus, controller, or network interconnects the components of the computing environment (100). Typically, operating system software (not shown) provides an operating environment for other software executing in the computing environment (100), and coordinates activities of the components of the computing environment (100).

[0023] The storage (140) may be removable or non-removable, and may include computer-readable storage media such as magnetic disks, magnetic tapes or cassettes, CD-ROMs, CD-RWs, DVDs, or any other medium which can be used to store information and which can be accessed within the computing environment (100). The storage (140) stores instructions for the software (180).

[0024] The input device(s) (150) may be a touch input device such as a keyboard, mouse, pen, or trackball; a voice input device; a scanning device; a network adapter; a CD/DVD reader; or another device that provides input to the computing environment (100). The output device(s) (160) may be a display, printer, speaker, CD/DVD-writer, network adapter, or another device that provides output from the computing environment (100).

[0025] The communication connection(s) (170) enable communication over a communication medium to another computing entity. Thus, the computing environment (100) may operate in a networked environment using logical connections to one or more remote computing devices, such as a personal computer, a server, a router, a network PC, a peer device or another common network node. The communication medium conveys information such as data or computer-executable instructions or requests in a modulated data signal. A modulated data signal is a signal that has one or more of its characteristics set or changed in such a manner as to encode information in the signal. By way of example, and not limitation, communication media include wired or wireless techniques implemented with an electrical, optical, RF, infrared, acoustic, or other carrier.

[0026] The tools and techniques can be described in the general context of computer-readable media, which may be storage media or communication media. Computer-readable storage media are any available storage media that can be accessed within a computing environment, but the term computer-readable storage media does not refer to propagated signals per se. By way of example, and not limitation, with the computing environment (100), computer-readable storage media include memory (120), storage (140), and combinations of the above.

[0027] The tools and techniques can be described in the general context of computer-executable instructions, such as those included in program modules, being executed in a computing environment on a target real or virtual processor. Generally, program modules include routines, programs, libraries, objects, classes, components, data structures, etc. that perform particular tasks or implement particular abstract data types. The functionality of the program modules may be combined or split between program

modules as desired in various embodiments. Computer-executable instructions for program modules may be executed within a local or distributed computing environment. In a distributed computing environment, program modules may be located in both local and remote computer storage media.

[0028] For the sake of presentation, the detailed description uses terms like "determine," "choose," "adjust," and "operate" to describe computer operations in a computing environment. These and other similar terms are high-level abstractions for operations performed by a computer, and should not be confused with acts performed by a human being, unless performance of an act by a human being (such as a "user") is explicitly noted. The actual computer operations corresponding to these terms vary depending on the implementation.

II. Dynamic Insertion of Synchronization Predicted Video Frames

A. System and Environment

[0029] Fig. 2 is a schematic diagram of a video transmission environment (200) in conjunction with which one or more of the described embodiments may be implemented. The environment (200) can include a coding computer system (210). The coding computer system (210) can code video data according to a standard or protocol and transmit the video data in a bitstream (220) to one or more decoding computer systems (230). This can be done using any of various different coding standards and techniques. For example, a scalable video standard may be used. An example of such a scalable video coding standard is H.264 SVC (scalable video coding), as described in Recommendation ITU-T H.264 (06/2011), although the techniques described herein could be used with other standards. The bitstream (220) can be transmitted over a computer network (240), which may include a global computer network (e.g., the Internet) and/or one or more other networks (e.g., an extranet, local area network, wide area network, etc.). Also, the transmission may be aided by a transmission server (250). For example, the transmission server (250) may act as an intermediary between the coding computer system (210) and the decoding computer system(s) (230). Also, the decoding computer system(s) (210) may encode their own bitstreams and send them out to other computer systems such as the coding computer system (210), and the coding computer system (210) may receive and decode video bitstreams. In one example, the computer systems (210 and 230) may be participating in a real time audiovisual event, such as a videoconference, where video bitstreams transmitted between multiple computer systems.

[0030] It is possible that data in the bitstream (220) can be lost during transmission to the decoding computer system(s) (230). For example, the lost data may be delayed so that the data arrives at the decoding computer system(s) too late to be used, or the data may never arrive at the decoding computer system(s). Either way, the data can be considered to be lost. The decoding computer system(s) (230) and/or the transmission server (250) may send one or more loss notices (260) to the coding computer system (210) to identify the lost data (e.g., identifying which frames and/or frame layers included all or part of the lost data). The loss notices (260) may be sent in transmissions using the same protocol as the bitstream (220), or in one or more out-of-band communications. Upon receiving such a loss notice (260), the coding computer system (210) can code and insert in the bitstream (220) a synchronization predicted video frame (270) that is coded using prediction, but where the prediction does not directly or indirectly reference the data identified in the loss notice (260) (the prediction does not reference the data identified in the loss notice (260), does not reference data that itself references the data identified in the loss notice (260), etc.). Such a video frame (270) can be used to synchronize the decoding computer system(s) (230) with the coding computer system (210), which may have been out of synchronization due to the lost data.

[0031] Accordingly, the video frame (270) can cut off drift that may have occurred due to the lost data. The synchronization predicted video frame (270) may be the next frame after a frame that includes the lost data, or it may be some later frame. For example, the coding computer system (210) may not receive the loss notice (260) until the coding computer system (210) has already coded and sent one or more subsequent frames in the bitstream (220). For the intervening frames between the frame that includes the lost data and the synchronization predicted video frame (270), the decoding computer system (230) may take measures to avoid or decrease the adverse effects from the reliance on lost data, such as by dropping or concealing those intervening frames.

B. Examples of Dynamic Insertion of Synchronization Predicted Video Frames

[0032] The coding computer system (210) may use different techniques and/or different types of inserted synchronization predicted video frames (270) to allow for synchronization. For scalable coded video such as H.264 SVC, performance may be improved by analyzing the location of the loss (such as by receiving a notice of data loss and a location (e.g., which frame and/or which layer) of the data loss) and inserting appropriate synchronization information based on the inter-layer dependency and

predictive coding structure. In the following description, dynamic synchronization video frame insertion will be discussed with reference to some predictive coding structures from H.264 SVC as an example, although the tools and techniques can be applied to other standards as well. Some examples of such techniques and tools will now be discussed with reference to Figs. 3-5.

[0033] Referring now to Fig. 3, an example of a regular prediction structure (300) with periodic key frames is illustrated. Each frame can include a base layer (302) and an enhancement layer (304) (or multiple enhancement layers, not shown), with the predictions' references being illustrated with arrows between the layers in the same frame and different frames. The video can be scaled by adding or omitting enhancement layers from a frame. For example, an enhancement layer may add additional quality features (e.g., smaller quantization step size) and/or higher spatial resolution. For example, the prediction structure (300) may be a prediction structure for a H.264 medium grain scalability (MGS) bitstream with periodic key frames. The illustrated regular prediction structure (300) shows frames numbered 0 to 10 in sequential order, and a similar order is shown for frames in the other prediction structures of Figs. 3-5.

[0034] The regular prediction structure (300) can start with an instantaneous decoding refresh (IDR) type key frame (310) (frame 0), which is an intra-coded key frame. A key frame, as used herein, is a frame that is limited to having no inter-frame predictions or only having inter-frame predictions that reference other key frames. The IDR-type key frame (310) can have intra-frame prediction, with an enhancement layer (304) (such as a quality enhancement layer) being coded with a prediction that references (directly or indirectly) the base layer of that frame and possibly lower enhancement layers of that frame. Additionally, and IDR-type key frame (310) can signal that subsequent frames should not include prediction references to frames prior to the IDR-type key frame (310).

[0035] The IDR-type key frame (310) can be followed by regular predicted frames (330) (frames 1, 2, 3, 4, 6, 7, 8, and 9). The regular predicted frames (330) can each include a base layer (302) and an enhancement layer (304). Each base layer (302) of a regular predicted frame (330) can be coded with a prediction that references a highest enhancement layer of the previous frame. Each enhancement layer (304) of a regular predicted frame (330) can be coded with a prediction that also references the highest enhancement layer of the previous frame, and that references the base layer (302) and/or one or more lower enhancement layers of that same frame.

[0036] The regular prediction structure (300) of Fig. 3 can include periodic insertion of predicted key frames (320) (frames 5 and 10). Each predicted key frame (320) can include a prediction that references other key frames (320 and/or 310), but does not reference regular predicted frames. Accordingly, a predicted key frame (320) can allow for synchronization between the coding computer system and decoding computer system(s) as long as previous key frames have not been lost. Accordingly, additional protections against loss may be used for the predicted key frames (320), as compared to protections used for the regular predicted frames (330). If a key frame (310 or 320) is lost, then the coding computer system can insert an IDR-type key frame (320) upon being notified of that loss.

[0037] Referring still to Fig. 3, a prediction structure (350) with dynamic predicted key frames will be discussed. The prediction structure (350) can be the same as the regular prediction structure (300) discussed above, except when the coding computer system receives a notification of lost data (360), which is illustrated with the layers including lost data being blacked out in Figs. 3-5. In frame 6, the lost data (360) is in the enhancement layer (304), but the base layer (302) was not lost. The next regular predicted frame (330) (frame 7) would have had a prediction that referenced the lost enhancement layer (i.e., the enhancement layer with the lost data) if the regular prediction structure (300) were followed. However, the coding computer system can receive a notice of the lost data and can code and insert a predicted key frame (320) as frame 7. That predicted key frame (320) can include a prediction that references the previous key frame (frame 5), but that does not reference the regular predicted frame (330) with the lost data (360) (because predicted key frames (320) are limited to not referencing such regular predicted frames (330)). Accordingly, the predicted key frame (320) inserted as frame 7 can act as a synchronization predicted video frame to cut off drift that may have occurred from the lost data (360) in the enhancement layer (304) of frame 6.

[0038] Referring still to Fig. 3, in frame 9, there is lost data (360) in the base layer (302) of frame 9. Because data in the base layer (302) is lost, and the enhancement layer (304) of frame 9 includes prediction that references the base layer (302), the enhancement layer (304) cannot be correctly decoded (as is depicted by the dashed lines around the enhancement layer (304) of frame 9). Accordingly, subsequent frames with predictions referencing the enhancement layer (304) of frame 9 would not be able to be correctly decoded, resulting in drift in those subsequent frames. However, the coding computer system can receive a notification of the lost data (360) in frame 9, and can respond by

coding and inserting a predicted key frame (320) as frame 10 in the bitstream, so that frame 10 can act as a synchronization predicted video frame, as with the predicted key frame (320) coded and inserted as frame 7. Because the predicted key frames (320) use some inter-frame prediction, they may be coded more efficiently than would intra-coded frames, such as IDR-type key frames, and they can still allow for synchronization and cutting off of drift from lost data.

[0039] Referring now to Fig. 4, another example of the use of dynamic insertion of synchronization predicted video frames will be discussed. For example, this may be used with a medium grain scalability (MGS) structure according to the H.264 SVC standard. A regular prediction structure (400) without periodic key frames is illustrated at the top of Fig. 4. The regular prediction structure (400) can start with a regular IDR-type key frame (410) as shown, or alternatively with an LIDR-type key frame (412) (see the bottom of Fig. 4). The LIDR-type key frame (412) can be similar to the IDR-type key frame (410) discussed above, except that that the long-term frame can be kept in the active frame window longer than would be done with a frame that is not a long term frame (e.g., a regular IDR-type key frame). The IDR frame (410) can be followed by regular predicted frames (430), which can be the same as the regular predicted frames (330) discussed above. The frames (410 and 430) can each have a base layer (402) and one or more enhancement layers (404), as discussed above.

[0040] The bottom of Fig. 4 illustrates a prediction structure (450) with an LIDR-type key frame (412) at the beginning and later with dynamic predicted key frames. As illustrated, the enhancement layer (404) of frame 6 and the base layer (402) of frame 9 include lost data (460). In these cases, the coding computer system can respond to each notice of lost data by coding and inserting in the bitstream a long term predicted key frame (420). Such a frame can be kept longer than regular frames, as discussed above with reference to the LIDR-type key frame (412). Accordingly, the latest key frame can be retained in the decoding buffer, at least until another key frame is coded and sent. Also, the long term predicted key frames (420) can include prediction that can reference other key frames, but that is limited to not referencing regular predicted frames (430). Accordingly, each long-term predicted key frame (420) can act as a synchronization predicted frame to allow synchronization of the coding and decoding computer systems after data is lost, which can cut off resulting drift.

[0041] Referring now to Fig. 5, another example of the use of synchronization predicted video frames with a different prediction structure will be discussed. A regular

prediction structure (500) without predicted key frames is illustrated at the top of Fig. 5. The regular prediction structure (500) can be similar to the regular prediction structure (400) discussed above. However, each regular predicted frame (530) can have a base layer (502) that includes a prediction that references a base layer (502) of a previous frame, rather than referencing an enhancement layer (504) of the previous frame. As an example, the regular prediction structure (500) may be structured according to a coarse grain scalable (CGS) bitstream or a spatial scalable bitstream under the H.264 SVC standard.

[0042] The bottom of Fig. 5 illustrates a prediction structure (550) with a dynamic anchor predicted frame. As illustrated in Fig. 5, the enhancement layer (504) of frame 6 and the base layer (502) of frame 9 include lost data (560). In response to receiving a notice of the lost data (560) in the enhancement layer (504) of frame 6, the coding computer system can code and insert an anchor predicted frame (520) as frame 7. The anchor predicted frame can include a base layer (502) with a prediction that references the base layer (502) of frame 6. The base layer (502) of the anchor predicted frame (520) can be coded as in the regular predicted frames (520), while the enhancement layer (504) of frame 7 can include a prediction that is limited to only including intra-frame references. Accordingly, the anchor predicted frame can include inter-frame prediction referencing the base layer (502) of the previous frame to improve coding efficiency, but can avoid inter-frame prediction referencing the enhancement layer (504) of frame 6 that includes lost data (560).

[0043] In response to receiving a notification of lost data (560) in the base layer (502) of frame 9, the coding computer system can code and insert an IDR-type key frame (510) as frame 10 to cut off drift from the lost data (560) in frame 9.

[0044] Note that combinations of the above types of synchronization predicted frames could be used. For example, combinations of anchor predicted frames (520) and long term predicted key frames (420) could be used to deal with losses in the same bitstream (e.g., in an H.264 CGS bitstream).

III. Techniques for Dynamic Insertion of Synchronization Predicted Video Frames

[0045] Several techniques for dynamic insertion of synchronization predicted video frames will now be discussed. Each of these techniques can be performed in a computing environment. For example, each technique may be performed in a computer system that includes at least one processor and memory including instructions stored thereon that when executed by at least one processor cause at least one processor to perform the technique (memory stores instructions (e.g., object code), and when

processor(s) execute(s) those instructions, processor(s) perform(s) the technique). Similarly, one or more computer-readable storage media may have computer-executable instructions embodied thereon that, when executed by at least one processor, cause at least one processor to perform the technique.

[0046] Referring to Fig. 6, a technique for dynamic insertion of synchronization predicted video frames will be described. The technique can include an encoding system encoding (605) and sending (610) a video bitstream over a computer network to a decoding computer system. The video bitstream can follow a regular prediction structure when the encoding computer system is not notified of lost data from the bitstream. The encoding computer system can receive (620) a notification of lost data in the bitstream. The lost data can include at least a portion of a reference frame of the bitstream. In response to the notification, the encoding computer system can dynamically encode (630) a synchronization predicted frame with a prediction that references one or more other previously-sent frames in the bitstream without referencing the lost data. The encoding computer system can insert (640) the synchronization predicted frame in the bitstream in a position where the regular prediction structure would have dictated inserting a different predicted frame with a prediction that would have referenced the lost data according to the regular prediction structure.

[0047] The synchronization predicted frame can include a predicted key frame whose prediction is limited to referencing one or more other key frames prior to the reference frame. The prediction of the predicted key frame can reference one or more intra-coded key frames (e.g. IDR frame(s)) and/or one or more predicted key frames.

[0048] The synchronization predicted frame can be a long term predicted key frame whose prediction is limited to referencing one or more other key frames prior to the lost data. The long term predicted key frame can reference one or more other long term key frames.

[0049] The different predicted frame that would have been coded and sent under the regular structure may have been a frame that would have referenced an enhancement layer of the reference frame. The synchronization predicted frame may be a key frame whose prediction references one or more key frames prior to the lost data. The lost data may include at least a portion of the enhancement layer of the reference frame and/or at least a portion of a base layer of the reference frame. The base layer can be referenced by prediction of the enhancement layer.

[0050] The lost data can include at least a portion of a lost enhancement layer, and the synchronization predicted frame can include an enhancement layer that references a base layer of the synchronization predicted frame. A prediction of the enhancement layer of the synchronization predicted frame can avoid referencing the lost enhancement layer. Additionally, an enhancement layer of the different predicted frame may have been a frame that would have referenced the enhancement layer with at least a portion of the lost data. A prediction of the base layer of the synchronization predicted frame may reference the base layer of a frame that includes at least a portion of the lost data.

[0051] Referring now to Fig. 7, a technique for dynamic insertion of synchronization predicted video frames will be described. The technique can include encoding (705) and sending (710) a video bitstream over a computer network to a decoding computer system. The video bitstream can follow a regular prediction structure when the encoding system is not notified of lost data from the bitstream. A notification of lost data in the bitstream can be received (720), where the lost data can include at least a portion of a reference frame of the bitstream. In response to the notification, the regular prediction structure can be dynamically modified (730) by encoding and inserting in the bitstream a synchronization predicted frame having a prediction that does not reference the lost data.

[0052] Inserting the synchronization predicted frame can include inserting the synchronization predicted frame in the bitstream in a position where the regular prediction structure would have dictated inserting a different predicted frame with a prediction that would have referenced the lost data according to the regular prediction structure. The different frame can be a frame that would have referenced an enhancement layer of the reference frame. For example, the enhancement layer of the reference frame may be a quality enhancement layer or a spatial enhancement layer. The lost data may include at least a portion of the enhancement layer, and the prediction of the synchronization predicted frame may reference a base layer below the enhancement layer without referencing the enhancement layer. The synchronization predicted frame may include a key frame whose prediction references one or more key frames prior to the lost data in the bitstream.

[0053] Referring now to Fig. 8, yet another technique for dynamic insertion of synchronization predicted video frames will be described. The technique can include an encoding computer system encoding (805) and sending (810) a real-time video bitstream over a computer network via a transmission server to a decoding computer system. The

video bitstream can follow a regular prediction structure between frames when the encoding computer system is not notified of lost data from the bitstream. The encoding computer system can receive (820) a notification of lost data in the bitstream. The lost data can include at least a portion of a reference frame of the bitstream. In response to the notification, the encoding computer system can dynamically encode (830) a predicted key frame whose prediction references one or more other key frames in the bitstream that are prior to the reference frame and does not reference the reference frame. The predicted key frame can be inserted (840) in the bitstream in a position where the regular prediction structure would have dictated inserting a different predicted frame with a prediction that would have referenced the lost data according to the regular prediction structure.

[0054] Although the subject matter has been described in language specific to structural features and/or methodological acts, it is to be understood that the subject matter defined in the appended claims is not necessarily limited to the specific features or acts described above. Rather, the specific features and acts described above are disclosed as example forms of implementing the claims.

CLAIMS

I/we claim:

1. A computer-implemented method, comprising:

an encoding computer system encoding and sending a video bitstream over a computer network to a decoding computer system, the video bitstream following a regular prediction structure when the encoding computer system is not notified of lost data from the bitstream;

the encoding computer system receiving a notification of lost data in the bitstream, the lost data comprising at least a portion of a reference frame of the bitstream;

in response to the notification, the encoding computer system dynamically encoding a synchronization predicted frame with a prediction that references one or more other previously-sent frames in the bitstream and that does not reference the lost data; and

the encoding computer system inserting the synchronization predicted frame in the bitstream in a position where the regular prediction structure would have dictated inserting a different predicted frame with a prediction that would have referenced the lost data according to the regular prediction structure.

2. The method of claim 1, wherein the synchronization predicted frame comprises a predicted key frame whose prediction is limited to referencing one or more other key frames prior to the reference frame.

3. The method of claim 1, wherein the synchronization predicted frame is a long term predicted key frame whose prediction is limited to referencing one or more other key frames prior to the lost data.

4. The method of claim 1, wherein the different predicted frame would have referenced an enhancement layer of the reference frame, and wherein the synchronization predicted frame is a key frame whose prediction references one or more key frames prior to the lost data.

5. The method of claim 4, wherein the lost data comprises at least a portion of the enhancement layer of the reference frame.

6. The method of claim 4, wherein the lost data comprises at least a portion of a base layer of the reference frame, the base layer being referenced by prediction of the enhancement layer.

7. The method of claim 1, wherein the lost data comprises at least a portion of a lost enhancement layer, wherein the synchronization predicted frame is an anchor frame that comprises an enhancement layer that references a base layer of the anchor frame,

wherein a prediction of the enhancement layer of the anchor frame avoids referencing the lost enhancement layer, and wherein an enhancement layer of the different predicted frame would have referenced the lost enhancement layer.

8. The method of claim 7, wherein a prediction of the base layer of the anchor frame references the base layer of a frame that includes at least a portion of the lost data.

9. A computer system comprising:

at least one processor; and

memory comprising instructions stored thereon that when executed by at least one processor cause at least one processor to perform acts comprising:

encoding and sending a video bitstream over a computer network to a decoding computer system, the video bitstream following a regular prediction structure when the encoding computer system is not notified of lost data from the bitstream;

receiving a notification of lost data in the bitstream, the lost data comprising at least a portion of a reference frame of the bitstream; and

in response to the notification, dynamically modifying the regular prediction structure by encoding and inserting in the bitstream a synchronization predicted frame having a prediction that does not reference the lost data.

10. One or more computer-readable storage media having computer-executable instructions embodied thereon that, when executed by at least one processor, cause at least one processor to perform acts comprising:

an encoding computer system encoding and sending a real-time video bitstream over a computer network via a transmission server to a decoding computer system, the video bitstream following a regular prediction structure between frames when the encoding computer system is not notified of lost data from the bitstream;

the encoding computer system receiving a notification of lost data in the bitstream, the lost data comprising at least a portion of a reference frame of the bitstream;

in response to the notification, the encoding computer system dynamically encoding a predicted key frame whose prediction references one or more other key frames in the bitstream that are prior to the reference frame and does not reference the reference frame; and

inserting the predicted key frame in the bitstream in a position where the regular prediction structure would have dictated inserting a different predicted frame with a

prediction that would have referenced the lost data according to the regular prediction structure.

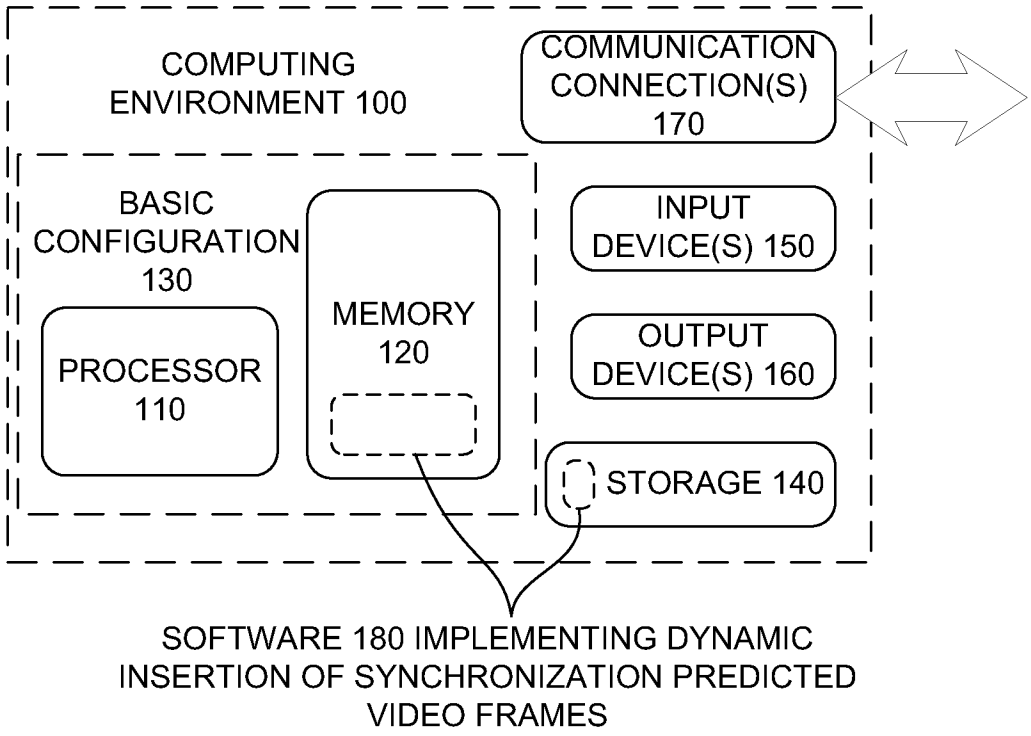

COMPUTING ENVIRONMENT 100
COMMUNICATION CONNECTION(S) 170
BASIC CONFIGURATION 130
PROCESSOR 110
MEMORY 120
INPUT DEVICE(S) 150
OUTPUT DEVICE(S) 160
STORAGE 140
SOFTWARE 180 IMPLEMENTING DYNAMIC INSERTION OF SYNCHRONIZATION PREDICTED VIDEO FRAMES

FIGURE 1

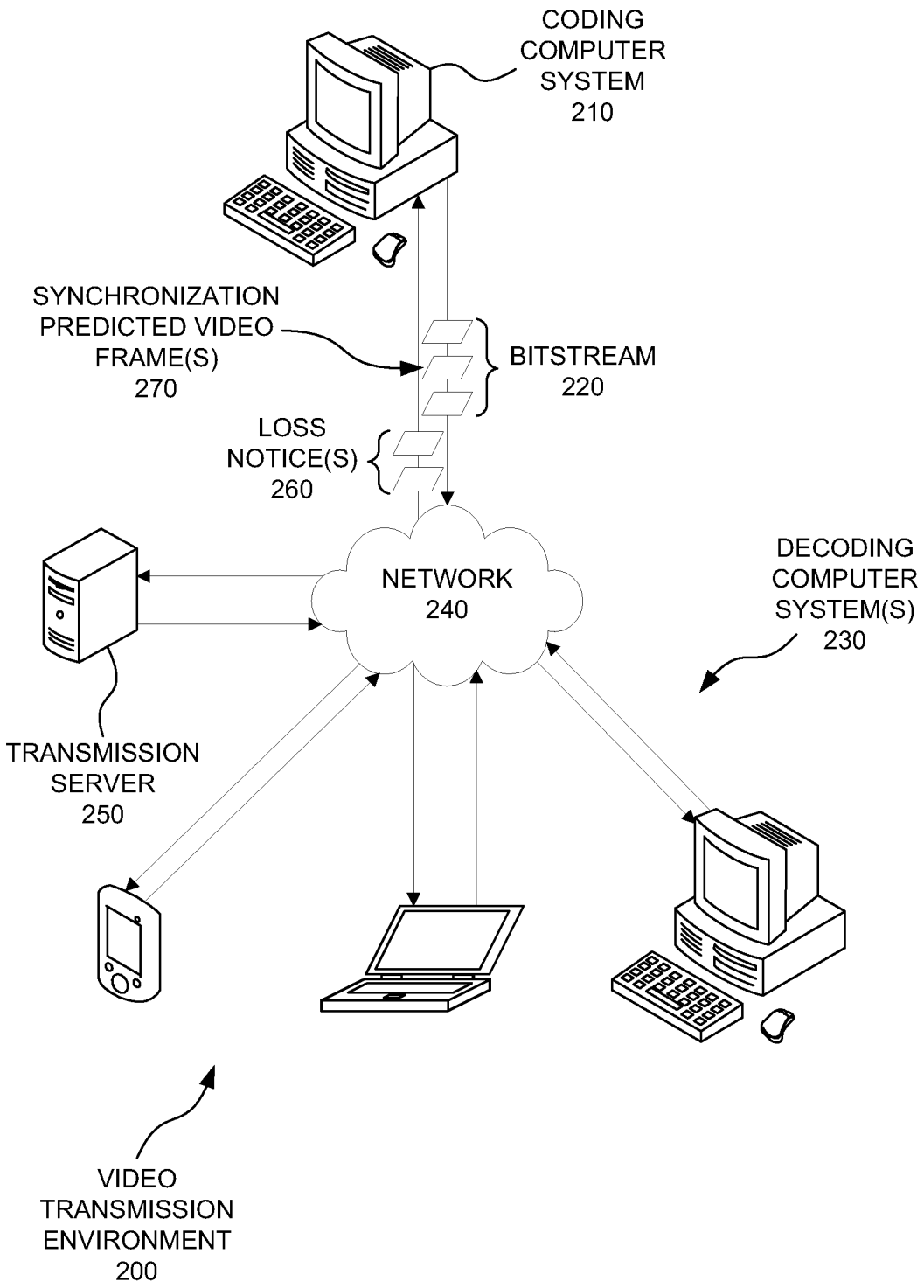

CODING
COMPUTER
SYSTEM
210
SYNCHRONIZATION
PREDICTED VIDEO
FRAME(S)
270
BITSTREAM
220
LOSS
NOTICE(S)
260
NETWORK
240
DECODING
COMPUTER
SYSTEM(S)
230
TRANSMISSION
SERVER
250
VIDEO
TRANSMISSION
ENVIRONMENT
200

FIGURE 2

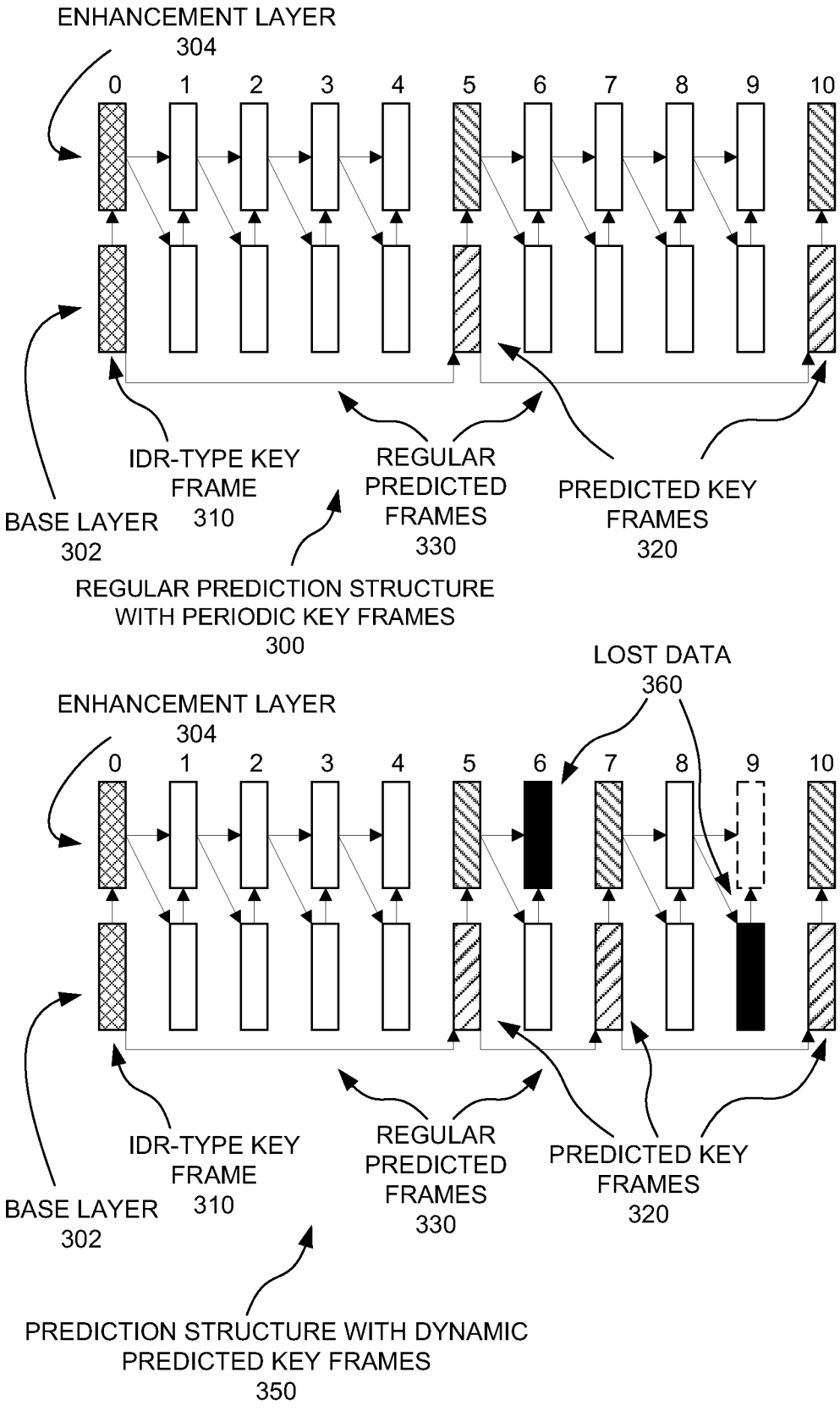

ENHANCEMENT LAYER
304
0 1 2 3 4 5 6 7 8 9 10
IDR-TYPE KEY FRAME 310
REGULAR PREDICTED FRAMES 330
PREDICTED KEY FRAMES 320
BASE LAYER 302
REGULAR PREDICTION STRUCTURE WITH PERIODIC KEY FRAMES 300
LOST DATA 360
ENHANCEMENT LAYER
304
0 1 2 3 4 5 6 7 8 9 10
IDR-TYPE KEY FRAME 310
REGULAR PREDICTED FRAMES 330
PREDICTED KEY FRAMES 320
BASE LAYER 302
PREDICTION STRUCTURE WITH DYNAMIC PREDICTED KEY FRAMES 350

FIGURE 3

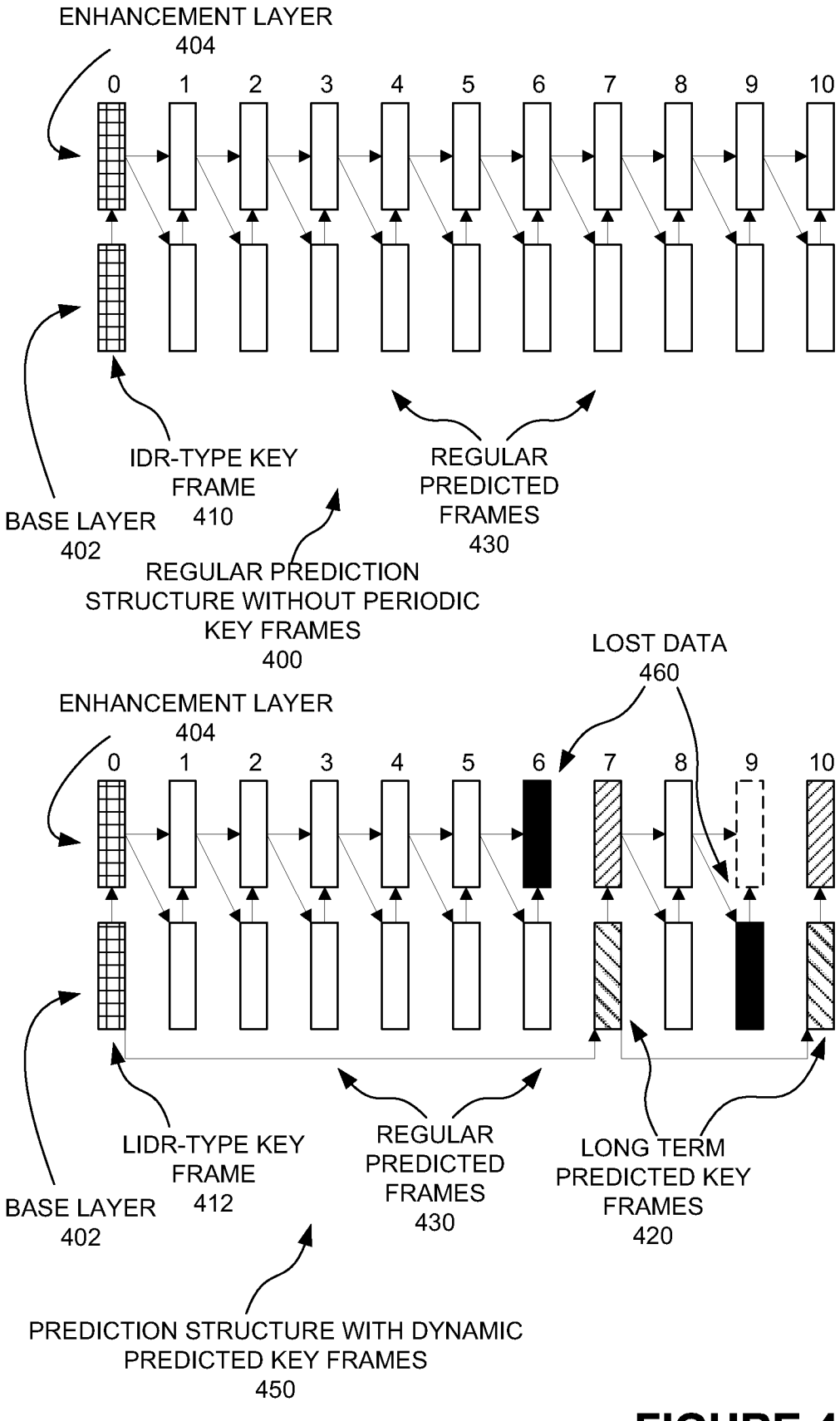

ENHANCEMENT LAYER
404
0 1 2 3 4 5 6 7 8 9 10
IDR-TYPE KEY FRAME 410
REGULAR PREDICTED FRAMES 430
BASE LAYER 402
REGULAR PREDICTION STRUCTURE WITHOUT PERIODIC KEY FRAMES 400
LOST DATA 460
ENHANCEMENT LAYER
404
0 1 2 3 4 5 6 7 8 9 10
LIDR-TYPE KEY FRAME 412
REGULAR PREDICTED FRAMES 430
LONG TERM PREDICTED KEY FRAMES 420
BASE LAYER 402
PREDICTION STRUCTURE WITH DYNAMIC PREDICTED KEY FRAMES 450

FIGURE 4

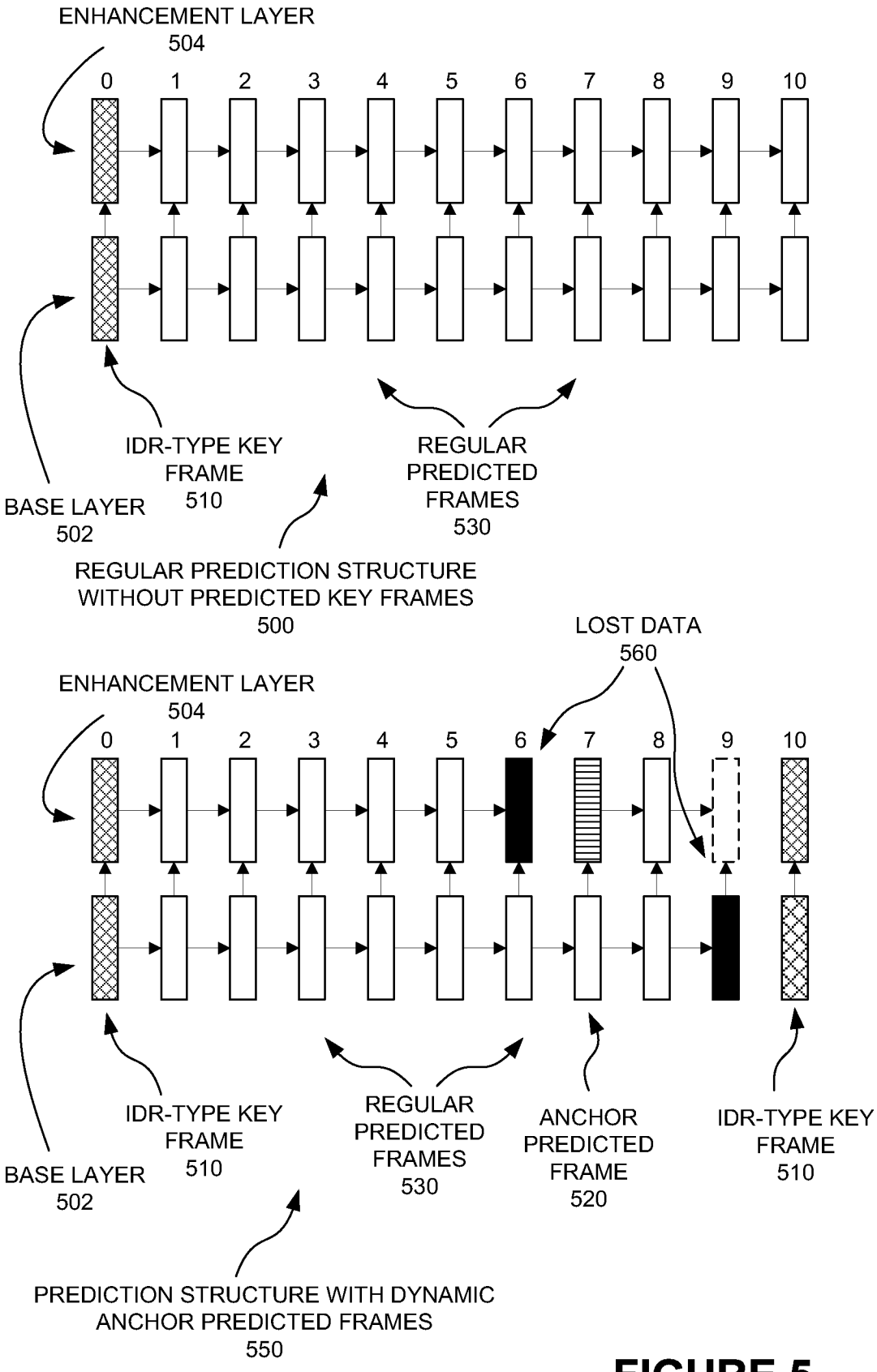

ENHANCEMENT LAYER
504
0 1 2 3 4 5 6 7 8 9 10
IDR-TYPE KEY FRAME 510
REGULAR PREDICTED FRAMES 530
BASE LAYER 502
REGULAR PREDICTION STRUCTURE WITHOUT PREDICTED KEY FRAMES
500
LOST DATA
560
ENHANCEMENT LAYER
504
0 1 2 3 4 5 6 7 8 9 10
IDR-TYPE KEY FRAME 510
REGULAR PREDICTED FRAMES 530
ANCHOR PREDICTED FRAME 520
IDR-TYPE KEY FRAME 510
BASE LAYER 502
PREDICTION STRUCTURE WITH DYNAMIC ANCHOR PREDICTED FRAMES
550

FIGURE 5

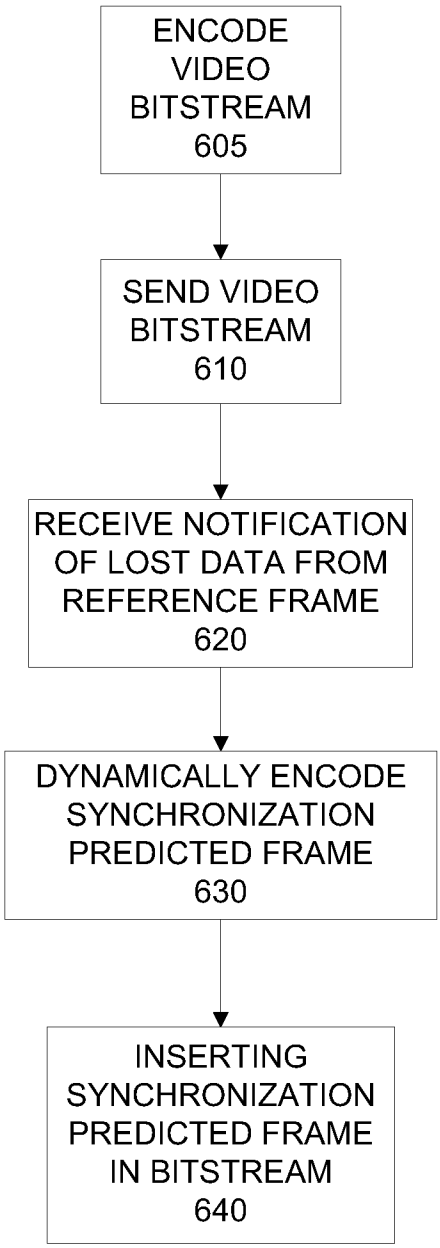

ENCODE VIDEO BITSTREAM 605
SEND VIDEO BITSTREAM 610
RECEIVE NOTIFICATION OF LOST DATA FROM REFERENCE FRAME 620
DYNAMICALLY ENCODE SYNCHRONIZATION PREDICTED FRAME 630
INSERTING SYNCHRONIZATION PREDICTED FRAME IN BITSTREAM 640

FIGURE 6

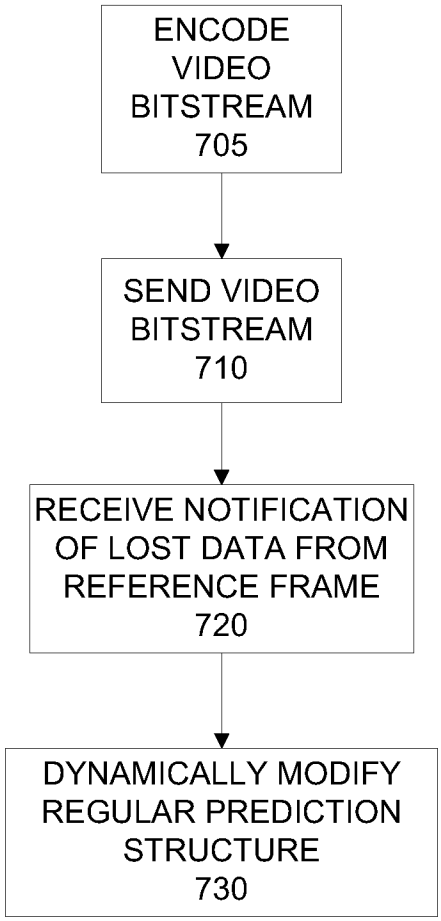

ENCODE VIDEO BITSTREAM 705
SEND VIDEO BITSTREAM 710
RECEIVE NOTIFICATION OF LOST DATA FROM REFERENCE FRAME 720
DYNAMICALLY MODIFY REGULAR PREDICTION STRUCTURE 730

FIGURE 7

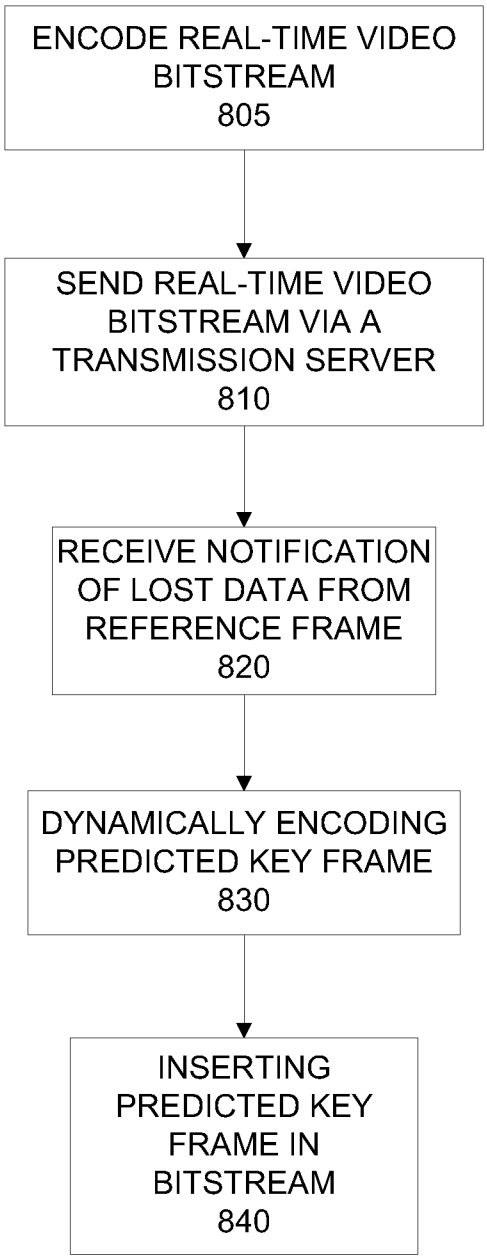

ENCODE REAL-TIME VIDEO BITSTREAM
805
SEND REAL-TIME VIDEO BITSTREAM VIA A TRANSMISSION SERVER
810
RECEIVE NOTIFICATION OF LOST DATA FROM REFERENCE FRAME
820
DYNAMICALLY ENCODING PREDICTED KEY FRAME
830
INSERTING PREDICTED KEY FRAME IN BITSTREAM
840

FIGURE 8

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2013/025800

A. CLASSIFICATION OF SUBJECT MATTER

H04N 7/24(2011.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
H04N 7/24; H03M 7/30; G06K 9/36; H04N 11/04; G06K 9/46; H04N 7/12; H03D 1/00; H04N 11/02; H04L 12/56

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Korean utility models and applications for utility models
Japanese utility models and applications for utility models

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
eKOMPASS(KIPO internal) & Keywords:predict, reference, enhancement layer, encoding, lost data, frame, impairment, corrupt, detection, damage, bitstream

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	KR 10-2011-0079579 A (MICROSOFT CORP.) 07 July 2011 See paragraphs [0031], [0062]; claims 1, 13; and figures 5a, 5b.	1-10
A	US 2009-0213938 A1 (LEE, YEN-CHI et al.) 27 August 2009 See paragraphs [0004]-[0005], [0007], [0051], [0145]-[0146]; and figure 3.	1-10
A	US 2006-0251191 A1 (WILLENEGGER, SERGE) 09 November 2006 See paragraphs [0038]-[0039]; and figures 1A, 2.	1-10
A	US 2006-0233239 A1 (SETHI, SUMEET et al.) 19 October 2006 See abstract; paragraphs [0012]-[0024]; claims 1, 4, 9; and figure 7.	1-10
A	US 2005-0105812 A1 (MOLINO, ANDREA et al.) 19 May 2005 See paragraph [0004]; and claim 1.	1-10

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents,such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
30 May 2013 (30.05.2013)	**03 June 2013 (03.06.2013)**

Name and mailing address of the ISA/KR
Korean Intellectual Property Office
189 Cheongsa-ro, Seo-gu, Daejeon Metropolitan City, 302-701, Republic of Korea
Facsimile No. 82-42-472-7140

Authorized officer
HWANG, Yun Koo
Telephone No. 82-42-481-5715

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.
PCT/US2013/025800

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
KR 10-2011-0079579 A	07.07.2011	CN 102158713 A	17.08.2011
		CN 1816150 A	09.08.2006
		CN 1816150 B	06.07.2011
		EP 1677547 A1	05.07.2006
		EP 1677547 B1	07.07.2010
		EP 2247111 A1	03.11.2010
		JP 05047497 B2	10.10.2012
		JP 2006-191599 A	20.07.2006
		JP 2012-090329 A	10.05.2012
		JP 2012-253831 A	20.12.2012
		KR 10-1114129 B1	20.02.2012
		KR 10-2012-0100851 A	12.09.2012
		US 2006-0146830 A1	06.07.2006
US 2009-0213938 A1	27.08.2009	CN 101960853 A	26.01.2011
		EP 2255536 A1	01.12.2010
		JP 2011-514076 A	28.04.2011
		KR 10-2010-0126410 A	01.12.2010
		KR 10-2013-0006524 A	16.01.2013
		TW 200943977 A	16.10.2009
		WO 2009-108614 A1	03.09.2009
US 2006-0251191 A1	09.11.2006	BR 0207876 A	21.12.2004
		CN 100473061 C0	25.03.2009
		CN 1498486 A	19.05.2004
		EP 1362459 A1	19.11.2003
		JP 2004-527158 A	02.09.2004
		KR 10-0892828 B1	10.04.2009
		TW 567698 A	21.12.2003
		US 2002-0141516 A1	03.10.2002
		US 7076005 B2	11.07.2006
		US 7656972 B2	02.02.2010
		WO 02-065720 A1	22.08.2002
US 2006-0233239 A1	19.10.2006	CN 101185336 A0	21.05.2008
		EP 1856915 A1	21.11.2007
		EP 1856915 B1	19.05.2010
		JP 2008-537373 A	11.09.2008
		KR 10-0928127 B1	25.11.2009
		TW 200704199 A	16.01.2007
		US 7929776 B2	19.04.2011
		WO 2006-099224 A1	21.09.2006
US 2005-0105812 A1	19.05.2005	US 7623719 B2	24.11.2009